On a Quadratic Euclidean Problem of Vector Subset Choice: Complexity and Algorithmic Approach

Anton Eremeev[1,2], Alexander Kel'manov[1,3], and Artem Pyatkin[1,3]

[1] Sobolev Institute of Mathematics,
4 Koptyug Ave., 630090 Novosibirsk, Russia
[2] Omsk State University n.a. F.M. Dostoevsky,
55A Mira Ave., 630077 Omsk, Russia
[3] Novosibirsk State University,
2 Pirogova St., 630090 Novosibirsk, Russia
eremeev@ofim.oscsbras.ru,{kelm,artem}@math.nsc.ru

Abstract. We analyze the complexity status of one of the known discrete optimization problems where the optimization criterium is switched from max to min. In the considered problem, we search in a finite set of Euclidean vectors (points) a subset that minimizes the squared norm of the sum of its elements divided by the cardinality of the subset. It is proved that if the dimension of the space is a part of input then the problem is NP-hard in a strong sense. Also, if the dimension of the space is fixed then the problem is NP-hard even for dimension 1 (on a line) and there are no approximation algorithms with guaranteed approximation ratio unless P=NP. It is shown that if the coordinates of the input vectors are integer then even a more general problem can be solved in a pseudopolynomial time in case when the dimension of the space is fixed.

Keywords: Euclidean space, subset search, clustering, NP-hardness, pseudopolynomial algorithms.

1 Introduction

The subject of study in the paper is one of the subset choice problems in a finite set of Euclidean vectors (points). The aim of the study is analysis of its complexity status and investigation of algorithmic approaches for it.

The investigation is motivated by poor (even almost no) understanding of the problem and its importance in applications, in particular, for mathematical problems of clustering, machine learning, computer geometry, data analysis and data mining. The problem is also interesting because the similar optimization problem with the opposite criterium is NP-hard in a strong sense (see below), and finding out the complexity status of the problems with opposite criteria is always an important theoretical study.

The paper is organized as follows. In the next section the formal definition of the currently studied and related problems are given; some examples of applications (origins) of the studied problem are also presented. In Section 3 we prove two theorems

In: A. Kononov et al. (eds.): DOOR 2016, Vladivostok, Russia, published at http://ceur-ws.org

on the complexity status of the problem (depending on whether the dimension of the space is a part of input or not). Finally, in Section 4 we formulate the problem as a series of integer linear programs and present a pseudopolynomial algorithm for it in the case of fixed dimension and integer coordinates of the vectors. The algorithm and linear programming formulation are applicable to a more general problem where a lower bound on cardinality of the sought subset is imposed.

2 Problem Formulation, Its Origin, Motivation and Related Problems

Everywhere below $\mathbb{R}$ denotes the set of real numbers, $\|\cdot\|$ denotes the Euclidean norm, and $\langle\cdot,\cdot\rangle$ denotes the scalar product.

We consider the following pair of problems.

Problem 1 (Subset with the Minimum Normalized Length of Vectors Sum). Given: a set $\mathcal{Y} = \{y_1, \dots, y_N\}$ of vectors (points) from $\mathbb{R}^q$. *Find*: a nonempty subset $\mathcal{C} \subseteq \mathcal{Y}$ such that

$$\frac{1}{|\mathcal{C}|}\Big\|\sum_{y\in\mathcal{C}} y\Big\|^2 \to \min\ .$$

The similar maximization problem is

Problem 2 (Subset with the Maximum Normalized Length of Vectors Sum). Given: a set $\mathcal{Y} = \{y_1, \dots, y_N\}$ of vectors from $\mathbb{R}^q$. *Find*: a subset $\mathcal{C} \subseteq \mathcal{Y}$ such that

$$\frac{1}{|\mathcal{C}|}\Big\|\sum_{y\in\mathcal{C}} y\Big\|^2 \to \max\ .$$

Both problems have an evident geometrical treatment — search for a geometrical structure with an optimal property (according to the corresponding objective function) in a finite set of points of Euclidean space. Both problems can be interpreted as problems of optimal summing. Besides, both problems are induced by applications listed below. Finally, note that for the objective function the equality

$$\frac{1}{|\mathcal{C}|}\Big\|\sum_{y\in\mathcal{C}} y\Big\|^2 = \sum_{y\in\mathcal{Y}} \|y\|^2 - \Big\{\sum_{y\in\mathcal{C}} \|y - \overline{y}(\mathcal{C})\|^2 + \sum_{y\in\mathcal{Y}\setminus\mathcal{C}} \|y\|^2\Big\} \qquad (1)$$

holds, where $\overline{y}(\mathcal{C}) = \frac{1}{|\mathcal{C}|}\sum_{y\in\mathcal{C}} y$ is a geometric center (centroid) of the subset $\mathcal{C}$.

In the right hand of the equality (1) the first sum is independent of $\mathcal{C}$. Therefore, the expression in the figure parenthesis can be considered as an objective function of the following pair of problems (maximum and minimum) of clustering the set $\mathcal{Y}$ into two subsets $\mathcal{C}$ and $\mathcal{Y}\setminus\mathcal{C}$.

Problem 3 (Maximum Sum-of-Squares 2*-Clustering with a Given Center). Given*: a set $\mathcal{Y} = \{y_1, \dots, y_N\}$ of vectors from $\mathbb{R}^q$. *Find*: a nonempty subset $\mathcal{C} \subseteq \mathcal{Y}$ such that

$$\sum_{y\in\mathcal{C}} \|y - \overline{y}(\mathcal{C})\|^2 + \sum_{y\in\mathcal{Y}\setminus\mathcal{C}} \|y\|^2 \to \max\ .$$

The similar minimization problem is

Problem 4 (Minimum Sum-of-Squares 2-Clustering with a Given Center).

Given: a set $\mathcal{Y} = \{y_1, \ldots, y_N\}$ of vectors from $\mathbb{R}^q$. *Find*: a subset $\mathcal{C} \subseteq \mathcal{Y}$ such that

$$\sum_{y \in \mathcal{C}} \|y - \overline{y}(\mathcal{C})\|^2 + \sum_{y \in \mathcal{Y} \setminus \mathcal{C}} \|y\|^2 \to \min .$$

It follows from (1) that Problems 3 and 4 are polynomially equivalent to Problems 1 and 2 respectively.

Recall that Problems 2 and 4 first arose in studying the problem of noise-proof off-line search for an unknown repeating fragment in a discrete signal [1]. The strong NP-hardness of these problems and their variations with additional restrictions on the cardinality of the desired set was proved in [2–6]. Problems 2 and 4, their generalizations and special cases were intensively studied in the latter decade [2–19]. Both problems have many applications (see the cited papers). In these papers some algorithmic results were obtained for these problems. The most important among them are following.

It was proved in [9, 14] that in the case of the fixed dimension q of the space Problems 2 and 4 are polynomially solvable in time $\mathcal{O}(N^{2q})$.

For Problem 2 in [3] an FPTAS is proposed for the case of fixed dimension q of the space. This scheme finds a solution with relative error at most $(q-1)/8l^2$, in time $\mathcal{O}(N^2(l)^{q-1})$, where l is an integer parameter of the algorithm.

For the variation of Problem 2 with an additional restriction that the cardinality of the subset equals M, heuristics based on local search ideas are proposed in [1, 2, 4]. Such algorithms having no guaranteed performance bounds are admissible for some applications.

For the same variation of Problem 2 and for the case of fixed space dimension in [7] and [14] exact pseudopolynomial algorithms are presented, whose time complexities are equal to $\mathcal{O}(NM(MD)^{q-1})$ and $\mathcal{O}(N(MD)^q)$, where D is the maximum absolute value of the coordinates of the input vectors. In [6] an FPTAS is proposed that finds an approximate solution with relative error $\varepsilon \leq (q-1)/4l^2$, in time $\mathcal{O}(N(\log N)(l)^{q-1})$, where l is an integer parameter of the algorithm. In [13] a randomized approximation algorithm is presented. Also in this paper some conditions under which the algorithm is polynomial and asymptotically precise are proved.

For Problem 4 in [11] a 2-approximation polynomial algorithm of complexity $\mathcal{O}(qN^2)$ is constructed.

For the variation of Problem 4 with additional restriction on the cardinality of the subset an algorithm that finds a 2-approximate solution in time $\mathcal{O}(qN^2)$ was proposed in [10]. A PTAS of complexity $\mathcal{O}(qN^{2/\varepsilon+1}(9/\varepsilon)^{3/\varepsilon})$, where ε is a guaranteed relative error is found in [15]. In [12] a randomized algorithm is presented that finds a $(1+\varepsilon)$-approximate solution of the problem in time $\mathcal{O}(qN)$ for a given relative error ε and probability of malfunction γ. Also, the conditions which ensure that this algorithm is asymptotically precise and has complexity $\mathcal{O}(qN^2)$ are found in [12].

For the same variation of Problem 4 and for the case of fixed space dimension in [16] an FPTAS is proposed. This scheme for a given relative error ε finds $(1+\varepsilon)$-approximate solution in time $\mathcal{O}(N^2(1/\varepsilon)^{q/2})$, that is polynomial in the size of input and $1/\varepsilon$.

As it was mentioned in the introduction, the issues of polynomial solvability and approximability for the opposite optimization criterium are traditional in discrete optimization. Therefore, investigating them for Problem 1 looks interesting. On the other hand these issues are also important for the applications where this problem arises. Among them is, for instance, physics. Indeed, if the objective function is written as

$$\frac{1}{|\mathcal{C}|}\Big\|\sum_{y\in\mathcal{C}} y\Big\|^2 = \sum_{x\in\mathcal{C}}\sum_{z\in\mathcal{C}}\langle x, z\rangle\,,$$

then the physical interpretation of Problem 1 is evident. Namely, the aim is to find a subset $\mathcal{C}$ of balanced forces (vectors) in the set $\mathcal{Y}$ of arbitrarily directed forces. Clearly, this problem has a wide range of technical applications. Also, under some conditions on the meaning of the coordinates of the input points, Problem 1 can be treated as finding a balanced by their views collective in a group of people having different views on some subject.

In the next section we prove that Problem 1 is NP-hard in the strong sense if the dimension of the space is a part of the input. In case of the fixed dimension of the space, that is typical for physical and technical applications, Problem 1 becomes NP-hard in ordinary sense even on line, i. e. even for $q = 1$. We also show that there are no approximation polynomial algorithms with guaranteed approximation ratio for Problem 1 unless P=NP. It follows from these results that all physical, technical and other applied problems of optimal balancing inducing Problem 1 are also hard.

3 Complexity Status

In order to analyse the complexity of Problem 1 we first formulate it as a decision problem.

Given: A set $\mathcal{Y} = \{y_1, \ldots, y_N\}$ of Euclidean points from $\mathbb{R}^q$ and a positive integer K. *Question*: Is there a nonempty subset $\mathcal{C} \subseteq \mathcal{Y}$ such that

$$\frac{1}{|\mathcal{C}|}\|\sum_{y\in\mathcal{C}} y\|^2 \leq K\,?$$

First consider the case when the dimension of the space is a part of input.

Theorem 1. *Problem 1 is NP-hard in the strong sense.*

Proof. We will prove that the corresponding decision problem formulated above is NP-complete. It is evident that this problem is in NP. We reduce to it the well-known NP-complete problem [20] *Exact cover by 3-sets.*

Problem Exact cover by 3*-sets. Given*: A finite set $\mathcal{Z}$ such that $|\mathcal{Z}| = 3n$, and a family $\mathcal{X} = \{\mathcal{X}_1, \ldots, \mathcal{X}_k\}$ of its subsets each of cardinality 3. *Question*: Is it true that the family $\mathcal{X}$ contains an exact cover of the set $\mathcal{Z}$, i. e. n subsets $\{\mathcal{X}_{i_1}, \ldots, \mathcal{X}_{i_n}\} \subseteq \mathcal{X}$, such that

$$\bigcup_{j=1}^{n} \mathcal{X}_{i_j} = \mathcal{Z}\,?$$

For a given instance of the problem *Exact cover by* 3-*sets* construct an equivalent instance of Problem 1 (in form of the decision problem) in the following way. Put $q = 3n$ and $K = 0$. For each subset $\mathcal{X}_i$, $i = 1, 2, \ldots, k$, introduce a corresponding vector y_i of dimension $3n$, whose j-th coordinate ($j = 1, \ldots, 3n$) is defined as follows: $y_i^{(j)} = 1$, if $j \in \mathcal{X}_i$, and $y_i^{(j)} = 0$ otherwise.

Let $y_{k+1} = (-1, \ldots, -1)$, $N = k + 1$, $\mathcal{Y} = \{y_1, \ldots, y_k, y_{k+1}\}$, and

$$z(\mathcal{C}) = \sum_{y \in \mathcal{C}} y\,.$$

Note that the objective function of Problem 1 is zero if and only if vector $z(\mathcal{C}) = 0$.

If the problem *Exact cover by* 3-*sets* has a positive answer then, clearly, the subset $\mathcal{C} = \{y_{i_1}, \ldots, y_{i_n}, y_N\}$ sums up to 0, and thus turns the objective function of Problem 1 to 0.

Now let $z(\mathcal{C}) = 0$ for some subset $\mathcal{C}$ in Problem 1. Then the subset $\mathcal{C}$ must contain the vector $y_N = (-1, \ldots, -1)$, since otherwise all coordinates of the vector $z(\mathcal{C})$ are non-negative and at least one of them is positive. But then all other vectors from the subset $\mathcal{C}$ must contain altogether exactly one 1 in each coordinate. Therefore, there are exactly n such vectors and the subsets corresponding to them form an exact cover.

Since the problem *Exact cover by* 3-*sets* is NP-hard in the strong sense and in the instance of Problem 1 constructed above all coordinates are at most 1 by absolute value, we proved the strong NP-hardness of Problem 1. □

Now assume that the dimension q of the space is fixed (not a part of input). In this case Problem 1 remains NP-hard, but not in the strong sense. This follows from the next theorem.

Theorem 2. *Problem 1 is NP-hard even for* $q = 1$.

Proof. Again we consider the decision form of Problem 1 and reduce to it the following NP-complete *Partition* problem [20].

Problem Partition. Given: An even number $n = 2k$ of positive integers α_j, $j = 1, \ldots, n$. *Question*: Is there a subset $\mathcal{I} \subset \{1, \ldots, n\}$, such that

$$\sum_{i \in \mathcal{I}} \alpha_i = \frac{1}{2} \sum_{i=1}^{n} \alpha_i\,.$$

For a given set of positive integers $\alpha_1, \ldots, \alpha_n$ from the instance of *Partitioning* problem construct an instance of Problem 1 in the following way. Put $q = 1$, $N = n + 1$, and $K = 0$.

Let

$$L = \sum_{i=1}^{n} \alpha_i\,.$$

Clearly, we may assume that L is even. Put $y_i = \alpha_i$ for $i = 1, \ldots, n$, $y_{n+1} = -L/2$ and, besides, for each subset $\mathcal{I} \subseteq \{1, \ldots, N\}$ define

$$S(\mathcal{I}) = \sum_{i \in \mathcal{I}} y_i\,.$$

If in the problem *Partition* there is a desired set $\mathcal{I}$, then it is easy to verify that $S(\mathcal{I} \cup \{n+1\}) = 0$, and thus the objective function of Problem 1 is equal to 0.

Assume now the instance of Problem 1 contains a subset $\mathcal{C}^*$ such that $z(\mathcal{C}^*) = 0$. Since y_{n+1} is the only negative number in $\mathcal{Y}$, we have $|\mathcal{C}^*| = k+1$ and $y_{n+1} \in \mathcal{C}^*$. But then since $z(\mathcal{C}^*) = 0$, we have

$$\sum_{i \in \mathcal{I}} \alpha_i = L/2 = \frac{1}{2} \sum_{i=1}^{n} \alpha_i \,,$$

where $\mathcal{I} = \{i \mid y_i \in \mathcal{C}^*\} \setminus \{n+1\}$.

So, Problem 1 is NP-hard even in the case when $q = 1$. □

4 Algorithmic Aspects

In this section we consider algorithmic aspects of Problem 1. In fact we will consider a more general

Problem 5 (Subset with the Minimum Normalized Length of Vectors Sum under Cardinality Restriction). Given: a set $\mathcal{Y} = \{y_1, \ldots, y_N\}$ of vectors (points) from $\mathbb{R}^q$ and an integer L, $1 \le L \le N$.

Find: a subset $\mathcal{C} \subseteq \mathcal{Y}$ such that $L \le |\mathcal{C}|$ and

$$\frac{1}{|\mathcal{C}|} \Big\| \sum_{y \in \mathcal{C}} y \Big\|^2 \to \min \,.$$

It is easy to see that Problem 1 is a partial case of Problem 5 for $L = 1$, and therefore all complexity results of the previous section are valid for Problem 5. We will also use an auxiliary problem denoted by Problem $1(M)$, where the criterion of Problem 1 needs to be minimized with an additional restriction that the size of the sought subset $\mathcal{C}$ is exactly M. Obviously, any instance of Problem 5 reduces to $N-L+1$ instances of Problem $1(M)$.

Consider Problem $1(M)$. Let us introduce Boolean variables $x_i,\ i = 1, \ldots, N$, such that $x_i = 1$ iff the i-th vector from $\mathcal{Y}$ is included into the set $\mathcal{C}$. Since $|\mathcal{C}| = M$, the sum of all x_i must be M. By the properties of the Euclidean norm we have

$$\frac{1}{|\mathcal{C}|} \left\| \sum_{i=1}^{N} x_i y_i \right\|^2 = \frac{1}{|\mathcal{C}|} \sum_{i=1}^{N} \|y_i\|^2 x_i + \frac{2}{|\mathcal{C}|} \sum_{k=2}^{N} \sum_{l=1}^{k-1} \langle y_k, y_l \rangle x_k x_l \,. \quad (2)$$

Define the auxiliary variables $z_{kl} \ge 0,\ k = 2, \ldots, N,\ l = 1, \ldots, k-1$, such that

$$z_{kl} = x_k x_l,\ k = 2, \ldots, N,\ l = 1, \ldots, k-1 \,.$$

This is equivalent to

$$z_{kl} \le x_k, \quad z_{kl} \le x_l, \quad k = 2, \ldots, N,\ l = 1, \ldots, k-1 \,,$$

$$z_{kl} \geq x_k + x_l - 1, \quad k = 2, \dots, N, \; l = 1, \dots, k-1 \, .$$

Put

$$C_i = \frac{1}{M} \|y_i\|^2, \;\; B_{kl} = \frac{2}{M} \langle y_k, y_l \rangle \, . \tag{3}$$

Then due to (2) and (3) the objective function of Problem 1 can be written as

$$\frac{1}{|\mathcal{C}|} \Big\| \sum_{y \in \mathcal{C}} y \Big\|^2 = \sum_{i=1}^{N} C_i x_i + \sum_{k=2}^{N} \sum_{l=1}^{k-1} B_{kl} z_{kl} \, .$$

So, we have reduced Problem 1(M) to the following mixed integer linear programming problem (note that all z_{kl} are Boolean automatically, i. e. we may not require this in the statement).

$$\sum_{i=1}^{N} C_i x_i + \sum_{k=2}^{N} \sum_{l=1}^{k-1} B_{kl} z_{kl} \to \min \, ,$$

where

$$z_{kl} \leq x_k, \quad k = 2, \dots, N, \; l = 1, \dots, k-1 \, ,$$

$$z_{kl} \leq x_l, \quad k = 2, \dots, N, \; l = 1, \dots, k-1 \, ,$$

$$z_{kl} \geq x_k + x_l - 1, \; k = 2, \dots, N, \; l = 1, \dots, k-1 \, ,$$

and

$$\sum_{i=1}^{N} x_i = M \, ,$$

$$x_i \in \{0, 1\}, \; i = 1, \dots, N \, ,$$

$$z_{kl} \geq 0, \; k = 2, \dots, N, \; l = 1, \dots, k-1 \, .$$

Note that the size of this integer program does not depend on q. Such a formulation can be helpful for applying various heuristics for Problem 5 in the case when q is a part of input.

Note that for any $r > 0$ there is no sense in finding polynomial r-approximation algorithms for Problem 5. Indeed, for any instance whose minimum of objective function is 0, an r-approximation algorithm must find an exact solution. In particular, it would find the optimal solution of the instances with $K = 0$ used in proofs of Theorems 1 and 2, i.e. it would solve problems *Exact cover by* 3*-sets* and *Partition* in a polynomial time, and this is impossible unless P=NP.

However, in case of a fixed dimension q of the space and integer coordinates of vectors from $\mathcal{Y}$ Problem 5 can be solved in a pseudopolynomial time.

For arbitrary sets $\mathcal{P}, \mathcal{Q} \subset \mathbb{R}^q$ define their sum as

$$\mathcal{P} + \mathcal{Q} = \{x \in \mathbb{R}^q \mid x = y + y', \; y \in \mathcal{P}, \; y' \in \mathcal{Q}\} \, . \tag{4}$$

For every positive integer r denote by $\mathcal{B}(r)$ the set of all vectors in $\mathbb{R}^q$ whose coordinates are integer and at most r by absolute value. Then $|\mathcal{B}(r)| \leq (2r+1)^q$.

Let b be the maximum absolute value of all coordinates of the input vectors $y_1, \ldots, y_N$. Our algorithm for Problem 5 successively computes the subsets $\mathcal{S}_k \subseteq \mathcal{B}(bk)$, $k = 1, \ldots, N$, that can be obtained by summing different elements of the set $\{y_1, \ldots, y_k\}$.

For $k = 1$ put $\mathcal{S}_1 = \{0, y_1\}$. Then we compute

$$\mathcal{S}_k = \mathcal{S}_{k-1} + (\{0\} \cup \{y_k\})$$

for all $k = 2, \ldots, N$, using the formula (4).

For each element $z \in \mathcal{S}_k$ we store an integer parameter n_z and a subset $\mathcal{C}_z \subseteq \mathcal{Y}$ such that

$$z = \sum_{y \in \mathcal{C}_z} y \,,$$

where $|\mathcal{C}_z| = n_z$ and n_z is the maximum possible number of addends that can be used to produce z.

Finally, find in the subset $\mathcal{S}_N$ an element $z^* \in \mathcal{S}_N$ such that $n_z \geq L$ and the value $\|z^*\|^2/n_{z^*}$ is minimum, and output the subset $\mathcal{C}_{z^*}$ corresponding to this z^*.

Computation of $\mathcal{S}_k$ takes $\mathcal{O}(q \cdot |\mathcal{S}_{k-1}|)$ operations. So, we have proved the following

Theorem 3. *If the coordinates of the input vectors from $\mathcal{Y}$ are integer and each of them is at most b by the absolute value then Problem 5 is solvable in $\mathcal{O}(qN(2bN+1)^q)$ time.*

Note that in case of fixed dimension q the running time of the algorithm is $\mathcal{O}(N(bN)^q)$, i. e. it becomes pseudopolynomial. So, we have the following

Corollary 1. *If the dimension q of the space is fixed, the coordinates of the input vectors from $\mathcal{Y}$ are integer and each of them is at most b by the absolute value then Problem 5 is solvable in pseudopolynomial time $\mathcal{O}(N(bN)^q)$.*

5 Conclusion

It follows from the obtained results that Problem 1 has neither polynomial nor pseudopolynomial exact algorithms unless P=NP. It also cannot be solved by a polynomial algorithm with a guaranteed approximation ratio. In the case of fixed dimension the problem remains NP-hard even for the line; however, it can be solved by a pseudopolynomial algorithm if the coordinates of the vectors are integer and the dimension is fixed.

The obtained results show that in spite of the easy looking formulation of the problem, finding effective algorithms for it looks hopeless except for the case when the dimension of the space is bounded by a constant and the coordinates of the input vectors are bounded by a polynomial in N. It seems also that obtaining positive results for this problem may be perspective in special cases modeling the specifics of applications excluding 0 from possible values of the objective function.

Acknowledgments. This work was supported by Russian Science Foundation. The results of Section 3 were obtained in frames of the project 16-11-10041. The results of Section 4 were obtained in frames of the project 15-11-10009.

References

1. Kel'manov A.V., Khamidullin S.A., Kel'manova M.A.: Joint Finding and Evaluation of a Repeating Fragment in Noised Number Sequence with Given Number of Quasiperiodic Repetitions (in Russian). In: Book of abstract of the Russian Conference "Discret Analysis and Operations Reserch" (DAOR-4), p. 185. Sobolev Institute of Mathematics SB RAN, Novosibirsk (2004)
2. Gimadi E.Kh., Kel'manov A.V., Kel'manova M.A., Khamidullin S.A.: Aposteriori Finding a Quasiperiodic Fragment with Given Number of Repetitions in a Number Sequence (in Russian). Sibirskii Zhurnal Industrial'noi Matematiki. 9(25), 55–74 (2006)
3. Baburin A.E., Gimadi E.Kh., Glebov N.I., Pyatkin A.V.: The Froblem of Finding a Subset of Vectors with the Maximum Total Weight. J. Appl. Indust. Math. 2(1), 32–38 (2008)
4. Gimadi E.Kh., Kel'manov A.V., Kel'manova M.A., Khamidullin S.A.: A posteriori Detecting a Quasiperiodic Fragment in a Numerical Sequence. Pattern Recognition and Image Analysis. 18(1), 30–42 (2008)
5. Kel'manov A.V., Pyatkin A.V.: On the Complexity of a Search for a Subset of "Similar" Vectors. Doklady Mathematics. 78(1), 574–575 (2008)
6. Kel'manov A.V., Pyatkin A.V.: On a Version of the Problem of Choosing a Vector Subset. J. Appl. Indust. Math. 3(4), 447–455 (2009)
7. Gimadi E.Kh., Glazkov Yu.V., Rykov I.A.: On Two Problems of Choosing some Subset of Vectors with Integer Coordinates that Has Maximum Norm of the Sum of Elements in Euclidean Space. J. Appl. Indust. Math. 3(3), 343–352 (2009)
8. Kel'manov A.V.: Off-line Detection of a Quasi-Periodically Recurring Fragment in a Numerical Sequence. Proceedings of the Steklov Institute of Mathematics. 263(S2), 84–92 (2008)
9. Gimadi E.Kh., Pyatkin A.V., Rykov I.A.: On Polynomial Solvability of Some Problems of a Vector Subset Choice in a Euclidean Space of Fixed Dimension. J. Appl. Indust. Math. 4(4), 48–53 (2010)
10. Dolgushev A.V., Kel'manov A.V.: An Approximation Algorithm for Solving a Problem of Cluster Analysis. J. Appl. Indust. Math. 5(4), 551–558 (2011)
11. Kel'manov A.V., Khandeev V.I.: A 2-Approximation Polynomial Algorithm for a Clustering Problem. J. Appl. Indust. Math. 7(4), 515–521 (2013)
12. Kel'manov A.V., Khandeev V.I.: A Randomized Algorithm for Two-Cluster Partition of a Set of Vectors. Comput. Math. and Math. Phys. 55(2), 330–339 (2015)
13. Gimadi E.Kh., Rykov I.A.: A Randomized Algorithm for Finding a Subset of Vectors. J. Appl. Indust. Math. 9(3), 351–357 (2015)
14. Kel'manov A.V., Khandeev V.I.: An Exact Pseudopolynomial Algorithm for a Problem of the Two-Cluster Partitioning of a Set of Vectors. J. Appl. Indust. Math. 9(4), 497–502 (2015)
15. Dolgushev A.V., Kel'manov A.V., Shenmaier V.V.: Polynomial-time Approximation Scheme for a Problem of Partitioning a Fnite Set into Two Clusters (in Russian). Trudy Instituta Matematiki i Mekhaniki UrO RAN. 21(3), 100–109 (2015)
16. Kel'manov A.V., Khandeev V.I.: Fully Polynomial-Time Approximation Scheme for a Special Case of a Quadratic Euclidean 2-Clustering Problem. Comput. Math. and Math. Phys. 56(2), 334–341 (2016)
17. Kel'manov A.V., Pyatkin A.V.: Complexity of Certain Problems of Searching for Subsets of Vectors and Cluster Analysis. Comput. Math. and Math. Phys. 49(11), 1966–1971 (2009)
18. Kel'manov A.V.: On the Complexity of Some Data Analysis Problems. Comput. Math. and Math. Phys. 50(11), 1941–1947 (2010)

19. Kel'manov A.V.: On the Complexity of Some Cluster Analysis Problems. Comput. Math. and Math. Phys. 51(11), 1983–1988 (2011)
20. Garey M.R., Johnson D.S.: Computers and Intractability: A Guide to the Theory of NP-Completeness. San Francisco: Freeman (1979)